

# SEQUENCE LISTING

SEQ ID NO:1 IND1 genomic

Sequence Range: 1 to 3856

5 CTCTAGACCA TCTACTATCC GGTGTTGAC CCTTAAAGCT TTTGAAGACT ACTAGAATAA 60

10 TGCAAATACC ATATGTCCAT ATCCATCCTT TTCTTTTGTT TGAAGTGAAC ATTCTAATTT 120

15 TGTAAGAGAA AAAACCTTAT GTTAATATCA CCGTAGGCAA AAAAAATATC TCATCATATT 180

20 AAATTTTAT TATAAGATTA TACATTCTCT CGTTGTAAGA GTTACTCCAA TTGCAAGTGT 240

25 TGTATTAAGT AATAAAAAGG ACGAAAGTAG GAAGCTTATA ATTAATTGAT GTTGCATAGT 300

30 ACTGGTATAT TGTGATGAA TATAACAAGT ATGAACATTA ATGCATGAAA CGGGGTATTT 360

35 TGTCTTGAAC TCATTAAAGG CAATGTGAAA AGAAGATGTG AGGTCTCATT TTGAAAATTT 420

40 ATCTTCTAGC TTTGTCGATT TTAATCTAT GAAATGAACG CAACATATAG AAATTTTCATG 480

45 TGGACAACGA CATTTAGACG GTATCTTAAT TAGACCGATT AATTAGTAAT ATACTTATAT 540

50 ATATAATTAG TGGTGATTAT AAGTTTACTT ATCCACTTGA GAATTTAAAC AATGGGCAAT 600

55 ACCTTAATGT CGAAAGAAGC CGTCCCACT TCGTGTAATG AGTTATGGGG GAGAGATCCT 660

60 GTTAAATCGT CAAATAAAAC AACTTAAGAA CTAGAAATTG ACACCAAAAA TCATAAGAG 720

65 AACGTTGAAG AAGTCATTTA TGGTATCCAG CTCATATTTT CTAGCTAAGA TCAAATCAAG 780

80 GCCGTTGAAA GGGCTTGTA GAAATGTGG AAGAAACCGT GGGGTTTAGA AGAAAGACAA 840

85 GAAATAGAAG AACAAATGATG TTAATTTGCC TATTTTGGTG TATAGGAGTT GTCAAAGAG 900

90 GAGAGAGAGA AGAAAATTAG GTCAAAATAA TGAGCACTAA AAATGGAGAC ATGTGTTGAG 960

95 TAACTATTAC AAGAGCGACT TATGCTTCCT TATGGCAATG ATATCCAAAC CAAAGTGCAA 1020

100 CGCTCCTTTT TTGCCCTAAT TTCGTAAAGT CTCTCTCCTT CTTCGTCCTT AGGAAAAACC 1080

105 CTAGAAATTT AATCCCTTGT TCTTGATCTT GCTTTTGGAG TAACCATGAT TTTGACCACA 1140

110 CACTATTTCT TCTATCTTTT GTGGTCTATA GGATTTTGCT TTATATGTGT TTCTGTATT 1200

115 GCTCCGTACG TACGTATACG AATTTAAATG GTTATAACAA GGTTTATATA AACTAGCACA 1260



AAAATATATC ATAAAATGAG ACATATATCC TATAAAAAAA AAATGAGGCA TATGAAGTAA  
 2700  
 5 ATAATAAGAG ACATGCATGT AAGCATTTCGG TTAATTAATC GAGTCAAAGA TATATATCAG  
 2760  
 TAAATACATA TGTGTATATT TCTGGA AAAA GAATATATAT ATTGAGAAAT AAGAAAAGAT  
 2820  
 10 GAAAATGGAA AATGGTATGT ATAAAAAGAA AGGAGTGTGC GACTCTTGTG TCTCGTCCAA  
 M E N G M Y K K K G V C D S C V S S K>  
 2880  
 15 AAGCAGATCC AACCACAGCC CCAAAAGAAG CATGATGGAG CCTCAGCCTC ACCATCTCCT  
 S R S N H S P K R S M M E P Q P H H L L>  
 2940  
 20 CATGGATTGG AACAAAGCTA ATGATCTTCT CACACAAGAA CACGCAGCTT TTCTCAATGA  
 M D W N K A N D L L T Q E H A A F L N D>  
 3000  
 25 TCCTCACCAT CTCATGTTAG ATCCACCTCC CGAAACCCTA ATTCACCTGG ACGAAGACGA  
 P H H L M L D P P P E T L I H L D E D E>  
 3060  
 AGAGTACGAT GAAGACATGG ATGCGATGAA GGAGATGCAG TACATGATCG CCGTCATGCA  
 E Y D E D M D A M K E M Q Y M I A V M Q>  
 3120  
 30 GCCCGTAGAC ATCGACCCTG CCACGGTCCC TAAGCCGAAC CGCCGTAACG TAAGGATAAG  
 P V D I D P A T V P K P N R R N V R I S>  
 3180  
 35 CGACGATCCT CAGACGGTGG TTGCTCGTCG GCGTCGGGAA AGGATCAGCG AGAAGATCCG  
 D D P Q T V V A R R R R E R I S E K I R>  
 3240  
 40 AATTCTCAAG AGGATCGTGC CTGGTGGTGC GAAGATGGAC ACAGCTTCCA TGCTCGACGA  
 I L K R I V P G G A K M D T A S M L D E>  
 3300  
 AGCCATACGT TACACCAAGT TCTTGAAACG GCAGGTGAGG ATTCTTCAGC CTCACTCTCA  
 A I R Y T K F L K R Q V R I L Q P H S Q>  
 3360  
 45 GATTGGAGCT CCTATGGCTA ACCCCTCTTA CCTTTGTTAT TACCACAACT CCCAACCTG  
 I G A P M A N P S Y L C Y Y H N S Q P \*>  
 3420  
 50 ATGAACTACA CAGAAGCTCG CTAGCTAGAC ATTTGGTGTC ATCCTCTCAA CCTTTTTCAT  
 3480  
 GTTGATATAT TATATATAGA TGCATAAAGA TTCGATCCAA GATTGTATGG GTGTTTAAAT  
 3540  
 55 ATTATTATTC TAAGATATAT GATGTACAAT TGTGTACCAA GTTTCTTTAT CTTGATATCA  
 3600  
 60 TATGCATAAA TAATTGGTGA ATAAAAAGAA GATATTGATT GTAAACAAAA AAAAGAAGAT  
 3660  
 ATTGATTGTT AATTAGGGTT TGATCATCTCT GTATGAAAGC TTTGGCCTGC AAATTAATTT  
 3720  
 65 TCGATATATA TATATATATA TGGAGAATAT ATATCAAATA CTTTTTTAAT TTGACTATAA  
 3780  
 TTTGTATCAA TTATCTGAAT CTGATGAGTG TAGGTTATAT ATGGATTAGC AAAAAAGAAA

3840  
ACAACCATTA TTACGCACCT ACATTAAAAA TCATCCACCA AAGAAGAAAC CATCCTCAAG  
5 AGGGTTCCCT CTAGAG

SEQ ID NO:2 IND1 protein

10 M E N G M Y K K K G V C D S C V S S K  
S R S N H S P K R S M M E P Q P H H L L  
M D W N K A N D L L T Q E H A A F L N D  
15 P H H L M L D P P P E T L I H L D E D E  
E Y D E D M D A M K E M Q Y M I A V M Q  
20 P V D I D P A T V P K P N R R N V R I S  
D D P Q T V V A R R R R E R I S E K I R  
I L K R I V P G G A K M D T A S M L D E  
25 A I R Y T K F L K R Q V R I L Q P H S Q  
I G A P M A N P S Y L C Y Y H N S Q P \*

SEQ ID NO:3 IND1 5' promoter

30 60 CTCTAGACCA TCTACTATCC GGTTGTTGAC CCTTAAAGCT TTTGAAGACT ACTAGAATAA  
120  
TGCAAATACC ATATGTCCAT ATCCATCCTT TTCTTTTGTT TGAAGTGAAC ATTCTAATTT  
35 180  
TGTAAGAGAA AAAACCTTAT GTTAATATCA CCGTAGGCAA AAAAAATATC TCATCATATT  
240  
40 AAATTTTAT TATAAGATTA TACATTCTCT CGTTGTAAGA GTTACTCCAA TTGCAAGTGT  
300  
TGTATTAAGT AATAAAAGG ACGAAAGTAG GAAGCTTATA ATTAATTGAT GTTGCAATAGT  
45 360  
ACTGGTATAT TGTGATGAA TATAACAAGT ATGAACATTA ATGCATGAAA CGGGGTATTT  
420  
50 TGTCTTGAAC TCATTAAAGG CAATGTGAAA AGAAGATGTG AGGTCTCATT TTGAAAATTT  
480  
ATCTTCTAGC TTTGTCGATT TTAAATCTAT GAAATGAACG CAACATATAG AAATTCATG  
540  
55 TGGACAACGA CATTTAGACG GTATCTTAAT TAGACCGATT AATTAGTAAT ATACTTATAT  
600  
ATATAATTAG TGGTGATTAT AAGTTTACTT ATCCACTTGA GAATTTAAAC AATGGGCAAT  
660  
60 ACCTTAATGT CGAAGAAGC CGTCCCACT TCGTGTAATG AGTTATGGGG GAGAGATCCT  
720  
GTAAATCGT CAAATAAAAC AACTTAAGAA CTAGAAATTG ACACCAAAAA TCATAAAGAG

	AACGTTGAAG	AAGTCATTTA	TCGTATCCAG	CTCATATTTT	CTAGCTAAGA	TCAAATCAAG	780
5							840
	GCCGTTGAAA	GGGCTTGTA	GAAAATGTCT	AAGAAACCGT	GGGGTTTAGA	AGAAAGACAA	900
10							960
	GAAATAGAAG	AACAATGATG	TTAAATTGCC	TATTTTGGTG	TATAGGAGTT	GTCAAAAGAG	1020
							1080
15							1140
	GAGAGAGAGA	AGAAAATTAG	GTCAAAATAA	TGAGCACTAA	AAATGGAGAC	ATGTGTTGAG	1200
							1260
20							1320
	TAACTATTAC	AAGAGCGACT	TATGCTTCCT	TATGGCAATG	ATATCCAAAC	CAAAGTGCAA	1380
							1440
25							1500
	CGCTCCTTTT	TTGCCCTAAT	TTCGTAAAGT	CTCTCTCCTT	CTTCGTCCTT	AGGAAAAACC	1560
							1620
30							1680
	CTAGAAATTT	AATCCCTTGT	TCTTGATCTT	GCTTTTGGAG	TAACCATGAT	TTTGACCACA	1740
							1800
35							1860
	CACTATTTCT	TCTATCTTTT	GTGGTCTATA	GGATTTTGCT	TTATATGTGT	TTCTTGATT	1920
							1980
40							2040
	GCTCCGTACG	TACGTATACG	AATTTAAATG	GTTATAACAA	GGTTTATATA	AACTAGCACA	2100
							2160
45							2220
	AATGAGTCCA	TGAAATTTGT	TAGCGAAAAA	GGTAGAAATA	TATTGAGTCT	TTAAACGGCA	2280
							2340
50							2400
	ATATATATAA	TTTTGCTGCA	AAACTTAGCT	TTAATCATGA	TCTAATGATA	TTTTCTTTAA	2460
							2520
55							2580
	TTTCCTTTGC	CAAATTAATC	ACATGCACGG	ATTTTGGCA	AGTTATGTGT	CGAATCTTTC	2640
							2700
60							2760
	CATTCACACA	ACACTAAACT	TAATTAGAAC	TCTAGGAAAT	ATTTTAAAT	GACAACTTTA	2820
							2880
65							2940
	TCGAAAAAAA	TTTAGTTATG	AAAACAATTC	CAGAATTAAA	CATGAGCTAT	ATAATTTAAG	3000
							3060
70							3120
	ATAAAATGAA	GTAATATTGA	TATGTATGTA	ATAACATATC	TGATTGCGGT	AAAAAAAAC	3180
							3240
75							3300
	ATATCTGATT	AAATTGTTCA	TGCAGGCCCA	TGTCATATG	ATGTCATCAC	GTTTTTATTT	3360
							3420
80							3480
	TCACAATAAC	TAATATATAT	TCAAAAAAAT	AGTTTGTCA	GATTAAATTT	TTTTTGGTGG	3540
							3600
85							3660
	TCAGCTTTCT	CCAACCTACT	AAACTAGTTT	GGAATGTTCT	CTTCTTTATT	TTTCTTTTTC	3720
							3780
90							3840
	TTGATTTCTT	ATGTTTTTTA	TTTATGGAAT	TTTAAGACGG	ATTGTTTAGG	TCGTTTCTCT	3900
							3960
95							4020
	CTTTTCTTGT	TTTCTAAAGT	TACTTTTGTA	AACTCATCTC	CTCCCAATTA	GACAGTCAAT	4080
							4140
100							4200
	CATATAGTTA	TCTTTTAATA	TATGTCTAGT	TGATAAAAAA	AATGAAAAAA	TACTGGTGGT	4260
							4320
105							4380
	AGTTCTACTA	ATGTTTGTTG	AAAAAATCTG	ATATTATGAA	TCTAATCAAT	TTCTTTGATC	4440
							4500

5

10

15

20

25

30

35

GTATAATGTG GGTAAATTT AGTAATTTT TACATAAATA AGAACTGTAA TGTGATGTA  
 2160  
 TATTGGGGAA TCAGTATATT AGCTTGGGTA ACTATACTTC TGGAAATACT TGAAGATTTA  
 2220  
 ACTATTTGCA AAATTATAAT TTAGTCCCGA AAAATACAGA CGACGGGACA CGACAACATA  
 2280  
 TAAGCAGGTT TGAATCTTGG AAAATTTTGT ATACATAACC TATATAAATA CTAATGTTCT  
 2340  
 GGTGGGTTC AAAAGCCTTT TCAAAAGTTC CATTTTTTAA ATTCAAGGAC ATTTTACATA  
 2400  
 GGAAATAAGT TGAGTCATAA AAAATAATGG TTATTTTGTA AGGTTTTTTT TTTGATTAAA  
 2460  
 ACGCACATAT TAAGAAGTTA GTTTTTTTTC ACTACCAAAT ATCAATTAAT TTAACCACAT  
 2520  
 GCAACCATTC ATAAACAAT ACTATTAAAG AATATAAATA ATCACAAAAT ATTAAATACA  
 2580  
 CTTAAATTT ACATATAAAT TTACAAAACA TCTAATTAAT TGAAACAGAA AGGAAAAGGT  
 2640  
 AAAATATATC ATAAATGAG ACATATATCC TATAAAAAAA AAATGAGGCA TATGAAGTAA  
 2700  
 ATAATAAGAG ACATGCATGT AAGCATTCCG TTAATTAATC GAGTCAAAGA TATATATCAG  
 2760  
 TAAATACATA TGTGTATATT TCTGGAAAAA GAATATATAT ATTGAGAAAT AAGAAAAGAT  
 GAAAA

SEQ ID NO:4 IND1 3' promoter

40

45

50

55

60

3420  
 ATGAACTACA CAGAAGCTCG CTAGCTAGAC ATTTGGTGTC ATCCTCTCAA CCTTTTTTCAT  
 3480  
 GTTGATATAT TATATATAGA TGCATAAAGA TTCGATCCAA GATTGTATGG GTGTTTTAAT  
 3540  
 ATTATTATTC TAAGATATAT GATGTACAAT TGTGTACCAA GTTCTTTTAT CTGATATCA  
 3600  
 TATGCATAAA TAATTGGTGA ATAAAAAGAA GATATTGATT GTAAACAAA AAAAGAAGAT  
 3660  
 ATTGATTGTT AATTAGGGTT TGATCATTCT GTATGAAAGC TTTGGCCTGC AAATTAATTT  
 3720  
 TCGATATATA TATATATATA TGGAGAATAT ATATCAAATA CTTTTTTAAT TTGACTATAA  
 3780  
 TTTGTATCAA TTATCTGAAT CTGATGAGTG TAGGTTATAT ATGGATTAGC AAAAAAGAA  
 3840  
 ACAACCATTA TTACGCACCT ACATTAAAAA TCATCCACCA AAGAAGAAAC CATCCTCAAG  
 AGGGTTCCCT CTAGAG

SEQ ID NO:5 SHP1 genomic

5 AGATCTGCAA CAGTGAAAAG AGAAAACAAA ATGGACTTGA AGAGGTTTTG ACAATGCCAG 60

10 AGATAATGCT TATTCCTTAA TATGTTGCCA GCCAAGTGTC AAATTGGCTT TTAAATATG 120

15 GATTCTCTGA TCAGTGGTCA TATTTGTGGA TCCAACGTAT TCATCATCAA GTTCTCAAGT 180

20 TTGCTTTCAG TGCAATTCTA ATTCACACGT TTAACCTTAA CATGCATGTC ATTATAATTA 240

25 CTTCTTCACT AAGACACAAT ACGGCAAACC TTCAGATTA TATTAATCTC CATAAATGAA 300

30 ATAATTAACC TCATAATCAA GATTCAATGT TTCTAAATAT ATATGGACAA AATTACACG 360

35 GAAGATTAGA TACGTATATT AGTAGATTTA GTCTTTCGTT TGTGCGATAA GATTAACCAC 420

40 CTCATAGATA GTAATATCAT TGTCAAATTC CTCTCGGTTT AGTCGCTAAA TTGTATCTTT 480

45 TTTAAGCCTA AAAGTAGTGT ATTCGCATAT GACTTATCGT CCTAACTTTT TTTTAAATTA 540

50 ACAAAAAAAT CGAAAAGAAA ATAATCTGTT AAATATTTTT TAAGTACTCC ATTAAGTTTA 600

55 GTTCTATTT AAAAAATGCT TGAAATTGGA CAGTTATGTT CAACAATTTT GAATCATGAG 660

60 CGATGTCTAG ATACTCAGAA TTAATCAAG ATGTCTTATC AAATTTGTTG TCACTCGAGG 720

65 ACCCACGCAA AAGAAAAGAC TAATATGATT TTTATTTGGT CTGGATATTT TTGTAGAGGA 780

70 TGAAACTAAG AGAGTGAAAG ATTCGAAATC CACAATGTTC AAGAGAGCTC AAAGCAAAAA 840

75 GAAAAATGAA GATGAAGGAC TAAAGAACAA TAAGCAACTA CTTATACCCT ATTTCCATAA 900

80 AGGATTCAGG TACTAGGAGA AGTTGAGGCA AGTTNNNNNN NATTGATTCA AATTTTCATT 960

85 TATTTTTACA ATTTAATTCA CCTAAGTTAT TATGCATTC TCATCATTGG TACATTTTCT 1020

90 GTATAGCGTA TTTACATATA TGAAATAAAT TAAATATGTC CTCACGTTGC AAGTAGTTAA 1080

95 TGAATGTCCC CACGCAAAAA AAAATCCCTC CAAATATGTC CACCTTTTCT TTTCTTTTAA 1140

100 ATTCCAAAAT TACCATAAAC TTTTGGTTTA CAAAAGATTT CTAGAAATTG AGGAAGATAT 1200

105 CCTAAATGAT TCATGAATCC TTCAATAATC TGAAGTTTGC GATATTTTCG ATTTCTTTCA 1260

110 AGAGTTGCGA TATTTGTAAT TTGGTGACCT TAACTTTTTT TTGATAAAGA GTAAACGTTT 1320





1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

2760  
TCAAGTTACT TTCTGCAAAC GACGCAATGG TCTTCTCAAG AAAGCTTATG AACTCTCTGT  
2820  
CTTGTGTGAT GCCGAAGTTG CCCTCGTCAT CTTCTCCACT CGTGGCCGTC TCTATGAGTA  
2880  
CGCCAACAAC AGGTACGCTT CTCCTACTCT ATTTCTTGAT CTTGTTTCT TAATTTTAAC  
2940  
TAAACAAGAT CCTAGTTCAA ATGATAACAA AGTGGGGATT GAGAGCCAAG ATTAGGGTTT  
3000  
GGTTAATTTA GAAAACCAGA TTTCACCTGT TGATACATTT AATATCTCTC TAGCTAGATT  
3060  
TAGTACTCTC TCCTCTATAT ATGTGTGGGT GTGTGTGTAA GTGTGTATAT GTATGCAAAT  
3120  
GCAAGAAGAA GAAGAAAAAG TTATCTTGTC TTCTCAAATT CTGATCAGCT TTGACCTTAG  
3180  
TTTCACCTCT TTTTCTGCAA ATCATTGAA CCTGATGCAT GTCAGTTTCT ACAATACACT  
3240  
TTTAATTTTG ACGGCCCATC AAATTTCCTA GGGTTTACTT CAGTGAACAA AATTGGGTTC  
3300  
TTGACACGAT TTAGCATGTA TATATAAAAA TAGGGGATGA TCAAGACTTA TGTAACCTCT  
3360  
GTCTGGTGAA ACTAGGGACA AAGTCTACTG ATGAGTTGTC ACTAGGGATC CATTTGATCA  
3420  
TTTAATCCCA AAAAAATGA AACAAAATTT TGAGAATTTA TATGCTGAAG TTTTCAACC  
3480  
CTCTTTTTTA AATAACTTTA TATTATGTAG ATTTGTATTT AGGGTAATTT GTCCAAC TAG  
3540  
AAGTCCTAAA AATCAATAAA CACACGGATG ACTTTGTCTA ACATTGTATC AGTCATCAA  
3600  
TGTAATAATTG TACAAATAAT GAAATTAAAG ATTTAGTCTC TTTTATTTTT TTGTTTAGG  
3660  
GTGTATATAT ATATATATAT GTATATTTGT TGCATTGATA TATCAATGAG AGGGAGAGAA  
3720  
CTCAGAGAAG TGTCGGAAAT TAAAATGGTA CGAGCCAATT GGAATCTCTG GCATTCTGAG  
3780  
CTTCATTGT TTGTTATTAG AAAAAAAAAA AAAAAATCCT TTAAAGATAC CTTCATGATG  
3840  
ACATTGAATC ATGTAATATA CACGATACAT GGTCTAATTC CTCCTCAAAC CTAATTACC  
3900  
AATTCGAAA CCATAATAT TACTAGTATG TTTATATATC CTTACTTTAA GACATTGTTT  
3960  
GTTTATAATA CCTGTGAAT TAAGAAAAAA AAAAAAAAC TTGTGGATCT ATTCAAGCCA  
4020  
TGTGTTAGAA TAAATTTATA AATTTTCTCC TCGTACTGGT CAGATATTGG TCCAACTCC  
4080

AAAGCCTTCC CTTTTCAGGA AAAAAACAT TTCGAAATTA ACTCTAATTA ATCAAGAATT  
4140  
5 TCCTACAATG TATACATCTA ATGTTTTTTC CGCGATCTTA CTTATTAGTG TGAGGGGTAC  
4200  
AATTGAAAGG TACAAGAAAG CTTGTTCGA TGCCGTCAAC CCTCCTCCG TCACCGAAGC  
4260  
10 TAATACTCAG GTACCAATTT ATATTGTTG ATTCTCTTG TTTTATCTTC TTCTTTTCAT  
4320  
TATATATATG ATCAACAAAA AATATAACCT ACAAAAAGAG AGAGTTCAAG GAAATGCATT  
4380  
15 GAAACGGTTT CGTTATGGTG TTTGAATACA TGGATTTTTG AAGTACTATC AGCAAGAAGC  
4440  
20 CTCTAAGCTT CGGAGGCAGA TTCGAGATAT TCAGAATTCA AATAGGTAAT TCATTAACCT  
4500  
TTCATGAACT CTTCGATTTG GTATTAGGTC ACTTAATTTG GTGTCGGTCC AAAAGTCCGC  
4560  
25 TTGTAGTTTT CTTTAGAAGT TGTTTTGTTT AATGTTTCATG TTTACAAATT GAAGGCATAT  
4620  
TGTTGGGGAA TCACTTGGTT CCTTGAACCT CAAGGAACTC AAAACCTAG AAGGACGTCT  
4680  
30 TGAAAAAGGA ATCAGCCGTG TCCGCTCCAA AAAGGTAAAA TCTACGTTGC TCTCTCTCTG  
4740  
35 TGTCTCTGTC TCTCTCTCTA TATATAGTCC CTTAGTTTAT ATAGTTCATC ACCCTTTTGT  
4800  
GAGAATTTTG CAGAATGAGC TGTTAGTGGC AGAGATAGAG TATATGCAGA AGAGGGTAAG  
4860  
40 AACGTTTCTC CCATTCCAAG TAATTAGATC TTTCTTCGTC TTTGTGAGGG TTTGAGTTTT  
4920  
CCCATAAATC ATGTGTAGGA AATGGAGTTG CAACACAATA ACATGTACCT GCGAGCAAAG  
4980  
45 GTTAGCCACG TTCTGTTCCA AATCTTAATC TCAATATCTA CTCTTTTCTT CATTTGTATAA  
5040  
50 CTAAGATAAC GTGAATAACA AGAAAACCTT TGTTTTTGGG TTTAATAGAT AGCCGAAGGC  
5100  
GCCAGATTGA ATCCGGACCA GCAGGAATCG AGTGTGATAC AAGGGACGAC AGTTTACGAA  
5160  
55 TCCGGTGTAT CTTCTCATGA CCAGTCGCAG CATTATAATC GGAATATAT TCCGGTGAAC  
5220  
CTTCTTGAAC CGAATCAGCA ATTCTCCGGC CAAGACCAAC CTCCTCTTCA ACTTGTGTAA  
5280  
60 CTCAAAACAT GATAACTTGT TTCTTCCCCT CATAACGATT AAGAGAGAGA CGAGAGAGTT  
5340  
65 CATTTTATAT TTATAACGCG ACTGTGTATT CATAGTTTAG GTTCTAATAA TGATAATAAC  
5400  
AAAACGTTG TTTCTTTGCT TAATTACATC AACATTTAAA TCCAAAGTTC TAAAACACGT

5460  
CGAGATCCAA AGTTTGTCAT ACAAGATTAG ACGCATACAC GATCAGTTAA TAGATTTTAA  
5520  
5  
GTGCCTTTTA ATATTTACAT ATAGTTGCAG CTTGATTAG ATCATGTCCA CCAAACACTC  
5580  
10  
ACAATTAGAG ACAAGCAAAA CTATAAACAT TGATCATAAA ATGATTACAA CATGTCCATA  
AATTAATTAT GGATTACAAA AATAAAAACT TACAAAAGAT CT

SEQ ID NO:6 SHP2 genomic

60  
15  
GAATTCGTAA CAGAATTTAG TGAATAATAT TGTAATTACC AGGCAAGGAC TCTCCAAACG  
120  
GATAGCTCGA ATATCGTTAT TAAAGAGTAA ATGATCCAAT ATGTAAGCCA TTGTTGATCA  
180  
20  
TCTAACATTG TTGGACTCTC TATTGCTCGA AATGATGCAT ACCTAATCAT TTATTCAGTT  
240  
25  
AACTATCAAG TTGCATTTGT AAAAACCAAA CATTTAAATT CAGATTTGAT ATCACTTACA  
300  
GAGGATAGAG AAGCATGACT CCAGGCCTGC ATGCAACAAG AAAAAGGAAG AAAATAATGT  
360  
30  
TAAAAATTG ACAAATATAG TGTTTATTTT TATTATATGA GACAGAATTT GAATAAAATC  
420  
CTACCCAACT AGAGCATCAA AACGTTTGC AATCGCAATA ATGAAACCCA TTTTCTTTT  
480  
35  
GAGTTTTTAC TCTTCTTTCA ACAGAACTT TCTCAAACGT CTTTAGCACT GTGACGTTAG  
540  
40  
ATATATACAC AAAAGCTTGA AATTTCTTCA AGCAAAAGAA TCTTTGTGGG AGTTAAGGCA  
600  
ACAAGCCAGG TAAAGAATCT CCAACGCATT GTTACGTTTT CATGAACCTA TTTATTATAT  
660  
45  
GTTCTAAGAA AGAAAAAAT ATCTCAAAGT AAACGTTGGA AATTTTCTGA TGAAGGGAAA  
720  
TCCAAAGTCT TGGGTTTAGT ATCCCTATGA ATGGTATTTG GAATATGTTT TCGTCAAAAC  
780  
50  
AAAAGATTCT TTTCTTTTTC ACAAGAGTTA GTGATCAATA ACTTATGCAC TAATTAATGA  
840  
55  
GATTGGACGT ATACACAATT TGATTATGAT ACTTGAGTAA AAATCACCTG TCCTTTAATT  
900  
TGGAATCTC TCTTCTTAC CCATTTATAT ACTACTTCTT TTCATTAAAA TTAAATTTC  
960  
60  
ATTATCAATC ATCGTTCAAT TTGATAAAGA TTAAACATTT TTTGTCACAG GGCTAGTAAA  
1020  
AGCAATCTTT ACATAATTCA TCTTCTTAC ATATATATAT TACCTTTTTC TTCATTAGTA  
1080  
65  
TTCTATTTGA TTATGATTAT TTTGTCATAA AGCTAGTAAA TTAAACACTC GATATGAGAA

1140  
TTATATTACT TCACGCTAAT TAACTCTTAA CACAACAAGA ACTAGTGCAT ATTCAACTTT

5 1200  
CAAAGCATAT ACTATATATT GAGAATATAG ACCACGAAAG TCAATCAAAA GACCTACCAG

10 1260  
CTCTCATCAA GTTCTTTCTT GAAATGATTT TGCAGAATTT CCAACTTAAT TAATTCGACA

15 1320  
TGAATGTGAA AATGTGTGTT GCTCGTTAAG AAAATTGAAT AGAAGTACAA TGAAATGAT

20 1380  
GAGGAATGGG CAAACACAA AAGAGTTTCC TTTCGTAAC ACAATTAATT AATGCAAATC

25 1440  
TGAGAAAGGG TTCATGGATA ATGACTACAC ACATGATTAG TCATTCCCCG TGGGCTCTCT

30 1500  
GCTTTCATTT ACTTTATTAG TTTCATCTTC TCTAATTATA TTGTCGCATA TATGATGCAG

35 1560  
TTCTTTTGTC TAAATTACGT AATATGATGT AATTAATTAT CAAAATAAAT ATTCAAATTG

40 1620  
CCGTTGGACT AACCTAATGT CCAAGATTAA GACTTGAACA TAAGAATTTT GGAAAACTA

45 1680  
AACCAGTTAT AATATATACT CTAAATGTC CATTCTGAA CACAACCAA TAATAATATA

50 1740  
TACTATTTAC AGTTTTTTTT AATTGGCAAG AACACTGAAA TCTTATTCAT TGTCTCGCTT

55 1800  
GGTAGTTGAC AAGTTATAAC ACTCATATTC ATATAACCCC ATTCTAACGT TGACGACGAA

60 1860  
CACTCATATA AACCACCCAA ATTCTTAGCA TATTAGCTAA ATATTGGTTT AATTGGAAAT

65 1920  
ATTTTTTTTA TATATAAAAT GCCAGGTAAA TATTAACGAC ATGCAATGTA TATAGGAGTA

2040  
GGGCAATAAA AAGAAAAGGA GAATAAAAAG GGATTACCAA AAAAGGAAAG TTTCCAAAAG

2100  
GTGATTCTGA TGAGAAACAG AGCCCATACC TCTCTTTTTT CCTCTAAACA TGAAAGAAAA

2160  
ATTGGATGGT CCTCCTTCAA TGCTCTCTCC CCACCCAATC CAAACCCAAC TGTCTTCTTT

2220  
CTTTCTTTTT TCTTCTTCT AATTTGATAT TTTCTACCAC TTAATTCCAA TCAATTTCAA

2280  
ATTTCAATCT AAATGTATGC ATATAGAATT TAATTAAG AATTAGGTGT GTGATATTTG

2340  
AGAAAATGTT AGAAGTAATG GTCCATGTTC TTTCTTTCTT TTTCCTTCTA TAACACTTCA

2400  
GTTTGAAAAA AACTACCAA ACCTTCTGTT TTCTGCAAAT GGGTTTTTAA ATACTTCCAA

2460  
AGAAATATTC CTCTAAAAGA AATTATAAAC CAAAACAGAA ACCAAAAACA AAAAATAAAG

TTGAAGCAGC AGTTAAGTGG TACTGAGATA ATAAGAATAG TATCTTTAGG CCAATGAACA  
 2520  
 AATTAACTCT CTCATAATTC ATCTTCCCAT CCTCACTTCT CTTTCTTTCT GATATAATTA  
 2580  
 ATCTTGCTAA GCCAGGTATG GTTATTGATG ATTTACACTT TTTTTTAAAA GTTTCTTCCT  
 2640  
 TTTCTCCAAT CAAATTCTTC AGTTAATCCT TATAAACCAT TTCTTTAATC CAAGGTGTTT  
 2700  
 GAGTGCAAAA GGATTTGATC TATTTCTCTT GTGTTTATAC TTCAGCTAGG GCTTATAGAA  
 2760  
 ATGGAGGGTG GTGCGAGTAA TGAAGTAGCA GAGAGCAGCA AGAAGATAGG GAGAGGGAAG  
 2820  
 ATAGAGATAA AGAGGATAGA GAACACTACG AATCGTCAAG TCACTTTCTG CAAACGACGC  
 2880  
 AATGGTTTAC TCAAGAAAGC TTATGAGCTC TCTGTCTTGT GTGACGCTGA GGTGCTCTT  
 2940  
 GTCATCTTCT CCACTCGAGG CCGTCTCTAC GAGTACGCCA ACAACAGGTA CACATCTTTT  
 3000  
 AGCTAGATCT TGATTTTGTT GAATTTTTTTT TCTAGAATAA AGTTTCGACT CTTCTGGTGG  
 3060  
 GTTTTTCAAT CTTTATGGTC TCTTTATAGT TTTTTCCTT AGTTTCTCTG AAGCTCAAAT  
 3120  
 CTCTTTAAAA ATCCCCAAAA TTAGGGTTTG TTTAAAACTA GGAACCCCTA CTTTAACTTC  
 3180  
 TTTCTCTTAG TAAAAAAGCA GTGAGGGTCT TCTCTGATCA TTAATTAGCA TCCCCCATA  
 3240  
 CTTGTTCAG TCACTTTTTC TCCACAAATC CTTATAACAG TATCTATATA TGTATCTATT  
 3300  
 TATGTCAGTT TGTACAAGAC ACTTCGATCA ATTTGATGAC CCATCAAGTT TTATTTCTGC  
 3360  
 AGATTGATCA TTAGGTTTCC ATCATAGTAA TGAAAAAGTA GGGTCTTGA TAAAATTATA  
 3420  
 ATAATATATA TTATTTGGCT ATATAAAAAA GCTATGTAGA TTCCTTAAAA ATTGATTCAC  
 3480  
 TAGGGAGAGA CTAGTAGGTG TTTGTCTTCT GACACTTCTC TAATCTTTTG GTGAATCCTT  
 3540  
 TTGTTAAATC AAGAAAATGA ATCAGGGACA AAGCTTATTG TTGAGTCACT TAATTAATCA  
 3600  
 TCCGATCCAT CAATCAAGAA AAATAACGAA ACAGAAAATT TTGATTTTGT ATTGTTATTT  
 3660  
 TCTCCACTTC AAGTTGGGGA CTTGTCATTT CCGTTTTTCT ATACGTTTCC AGCTATTAAC  
 3720  
 AGCTCATGTT CATTTCACCA TTTTGATTAT TTGTCTGCTT TTAAAGATA AATGTTTTCA  
 3780  
 AAAATATGT TTTTATTTGC TTGGCTAGTT AATACTATAA TTGAGGTGA TGTATGACTA

3840  
 TAATCTATAA GTCAAGTCTC ATATCATGGA TCTAAGTTAA AACTAGTAAA TTTGTAGTTT  
 3900  
 5 CAATGTGAAC TTTCACAACG ACTAAAGAAC TGATCTGAAG TTTATAATGG ACATGACTAA  
 3960  
 TTTGATTAAAC AAAAGAGGAA TGCATTATGT ATGTAGAAAC ATGTGATATA TATATGTTTC  
 4020  
 10 TATTATCAAA AGTGTAGTTA ACTTTCTTAT TTCAAACACC CTCATGCTTT AGTAGTATCT  
 4080  
 15 TACTTTTGAC ATTTCTCAAC TTCAGCTTTC CATTATACAA CAGCACAATG TAAATTAATT  
 4140  
 GTATATGAAT ATGAAAGCAT AACGTTATGC AAAGATTTCT AGCTTTTCTT TTTCTGTTTT  
 4200  
 20 GCAAAAGATT TACAAATATC ATGTTCTTGG TAAAAACATA CTTGCCTCAG CCACATATGC  
 4260  
 ATGTAAATGT AATGTTCAAA TATTAATTCA GGAAAAACAA AGAAGAAGCA AAATTAGCTT  
 4320  
 25 CTAGAGTAGG GAATCTATTG ACTTGACCTG AAAATCACTT CTTTTTCTTA AAGCCTAGTA  
 4380  
 30 GTGAATTTT TAATCTAATT AGGCCAAAAT ATATACTAGC CTAAATATA ATTTGGATT  
 4440  
 TGTGTCGTAC ATAAATTGGG ACCAATTCCA ATTAATAAG AGCATATGCA ATTCAAATTC  
 4500  
 35 TTTTATTTT CTTCTCCGAT TTGCTACTTC TTTCTTTTGT ATGTTTCAA ATTAGGATTA  
 4560  
 CACTTTTTTG GGGAAGTACA CATTAGGTC TTCTCGAACT TTGATTATAC ATATATATAT  
 4620  
 40 ATATATATAT ATATAACTTT GTGAGATGTC ACTGTTAATA GATAATAGGC AATAACAATA  
 4680  
 ATATCCAAAA AAGAAGGCGC AAACAAATCA TATACTATAT GGTACTGGTC CATTCACTAT  
 4740  
 45 TTTGTCGGTT GAATTTAAGG TTTGGCGTAC AACTTTGTT TCAAACCTT ATTATCCGT  
 4800  
 50 CTTTCTGTGT GTTTGTATA TCCAGAAGAT AAAAATATCA ATTTCTTTAA CGACTTCATA  
 4860  
 TATATATATA TATATATATA TATATATATT TTTCTCTTCT GGTTTTAGTG TTTGAATCCA  
 4920  
 55 ACAGTTATAG TTTCGTGTGT CTTTGTTTTA CTTGTGGTGG TTTAAGTTG AGATTTTCAC  
 4980  
 60 CGATTGCATC TATTACATA TATAGCTACC ACAAAAAAGA TTGCATTTA AAATCTTTTC  
 5040  
 CTTTGTGTGA ATGTTGATGA AGTGTGAGAG GAACAATAGA AAGGTACAAG AAAGCTTGCT  
 5100  
 65 CCGACGCCGT TAACCCTCCG ACCATCACCG AAGCTAATAC TCAGGTTAGC TTTTAATTAA  
 5160  
 TACACCTAGC TAGCTAGTTC GTTAATTACT TAATTTCTTC TTCTTTTAGT TATCTGACCT

5  
10  
15  
20  
25  
30  
35  
40  
45  
50

TTTTTTCACC TCTTGTAACA ATGATGGGAT CGAAATTGAT GAAGTACTAT CAGCAAGAGG 5220  
CGTCTAAACT CCGGAGACAG ATTCGGGACA TTCAGAATTT GAACAGACAC ATTCTTGGTG 5280  
AATCTCTTGG TTCCTTGAAC TTTAAGGAAC TCAAGAACCT TGAAAGTAGG CTTGAGAAAG 5340  
GAATCAGTCG TGTCCGATCC AAGAAGGTAC ATCACTAACT CTCCATCAAT CTCCTTATCA 5400  
TTGAATATAT ATCCATCTGA TTCTTGCCCG TTATATTTGG TTTTCTCTC CAGCACGAGA 5460  
TGTTAGTTGC AGAGATTGAA TACATGCAA AAAGGGTAAA AGTAAACCT ATCTTCCTTC 5520  
ACAATGAACT ACCCCTACTT TATTAGCAAC TTCTCTTTCT GATGATCATC TTTTTTATTT 5580  
TCTGTTGTCG CTTGCATTGT AGGAAATCGA GCTGCAAAAC GATAACATGT ATCTCCGCTC 5640  
CAAGGTTTTA TACATAACTC TTTTGGCAT TTTTGATCAT CATTTTTTTC CGGTAGACAA 5700  
TCTCTTGATG TGCAAATTCT AAATATCTCT GCAGATTACT GAAAGAACAG GTCTACAGCA 5760  
ACAAGAATCG AGTGTGATAC ATCAAGGGAC AGTTTACGAG TCGGGTGTTA CTTCTTCTCA 5820  
CCAGTCGGGG CAGTATAACC GGAATTATAT TGCGGTTAAC CTTCTTGAAC CGAATCAGAA 5880  
TTCCTCCAAC CAAGACCAAC CACCTCTGCA ACTTGTTTGA TTCAGTCTAA CATAAGCTTC 5940  
TTTCCTCAGC CTGAGATCGA TCTATAGTGT CACCTAAATG CGGCCGCGTC CCTCAACATC 6000  
TAGTCGCAAG CTGAGGGGAA CCACTAGTGT CATACGAACC TCCAAGAGAC GGTTACACAA 6060  
ACGGGTACAT TGTGATGTC ATGTATGACA ATCGCCCAAG TAAGTATCCA GCTGTGTTCA 6120  
GAACGTACGT CCGAATTC